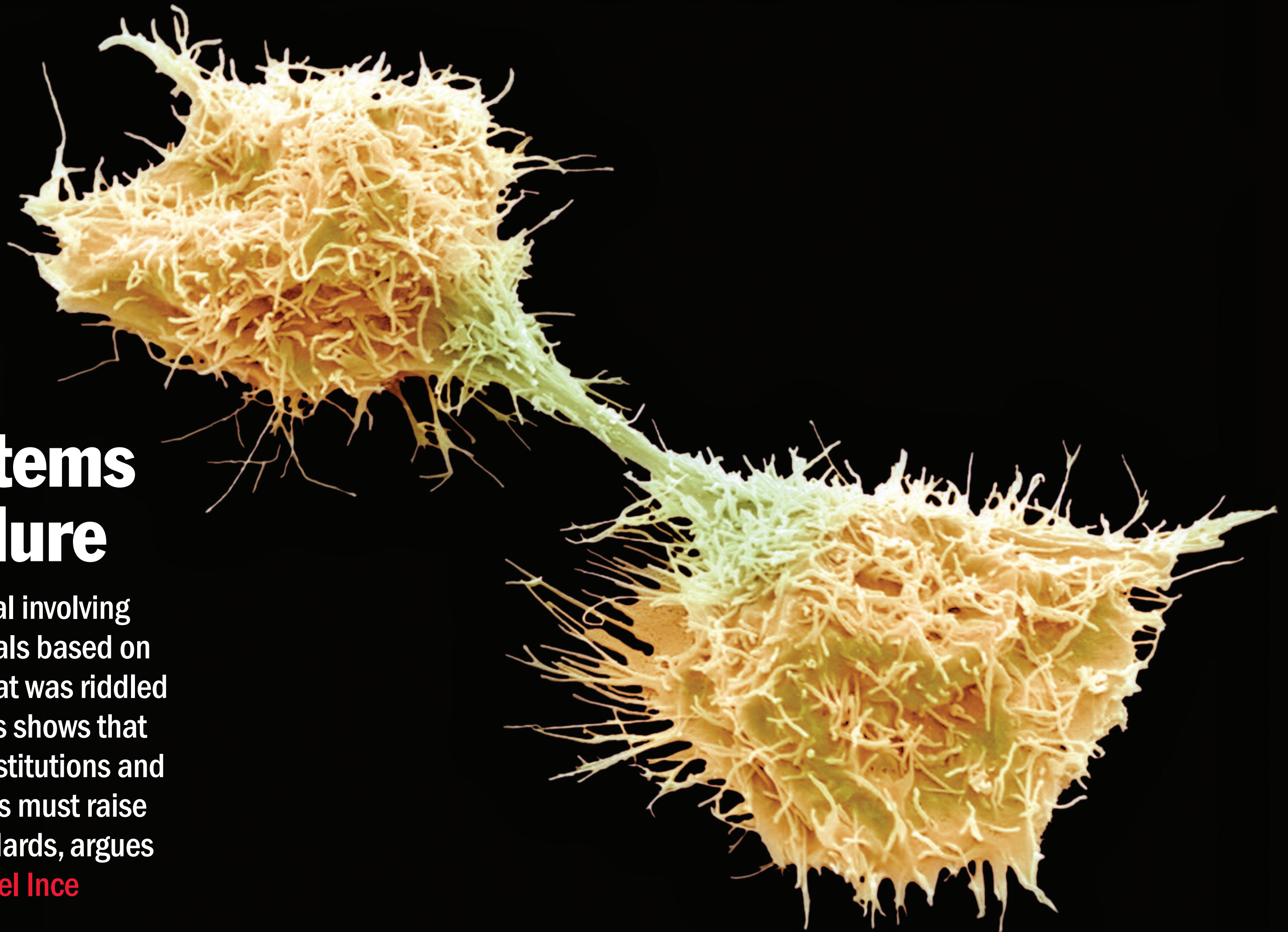


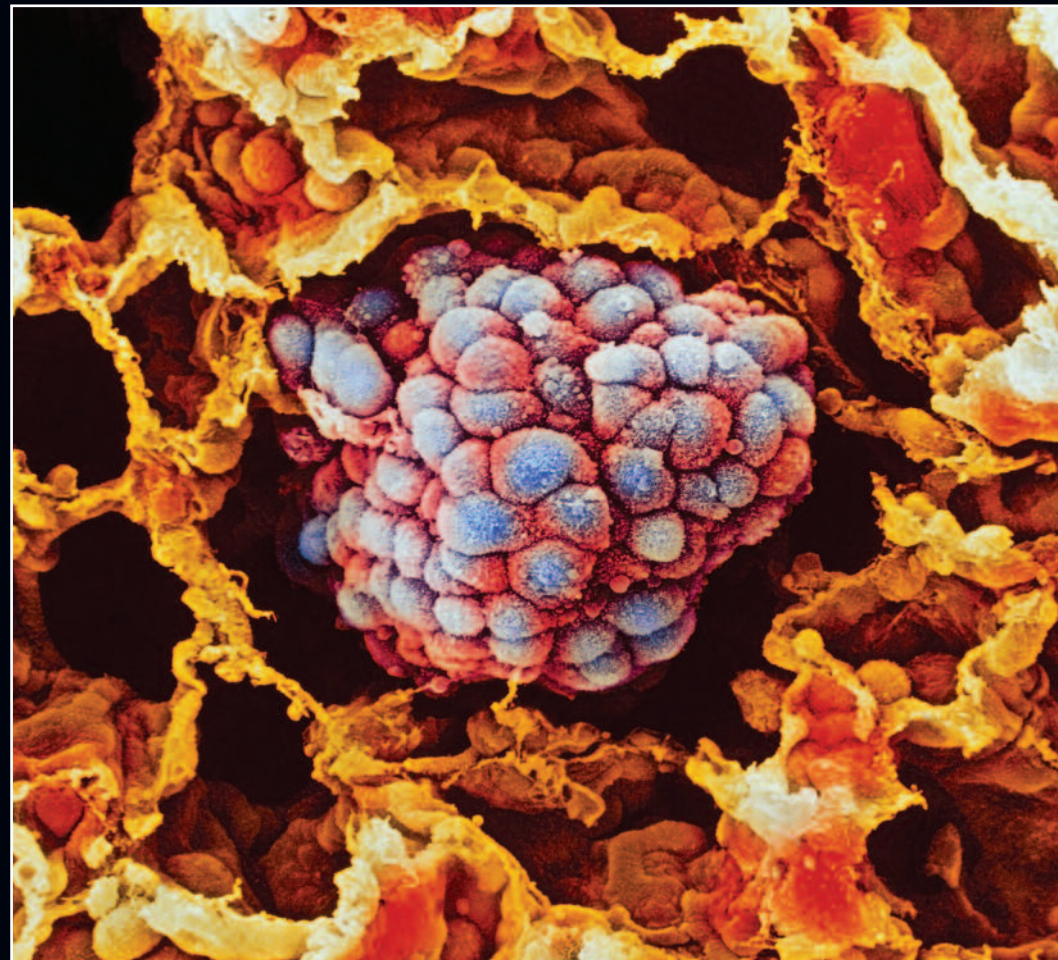
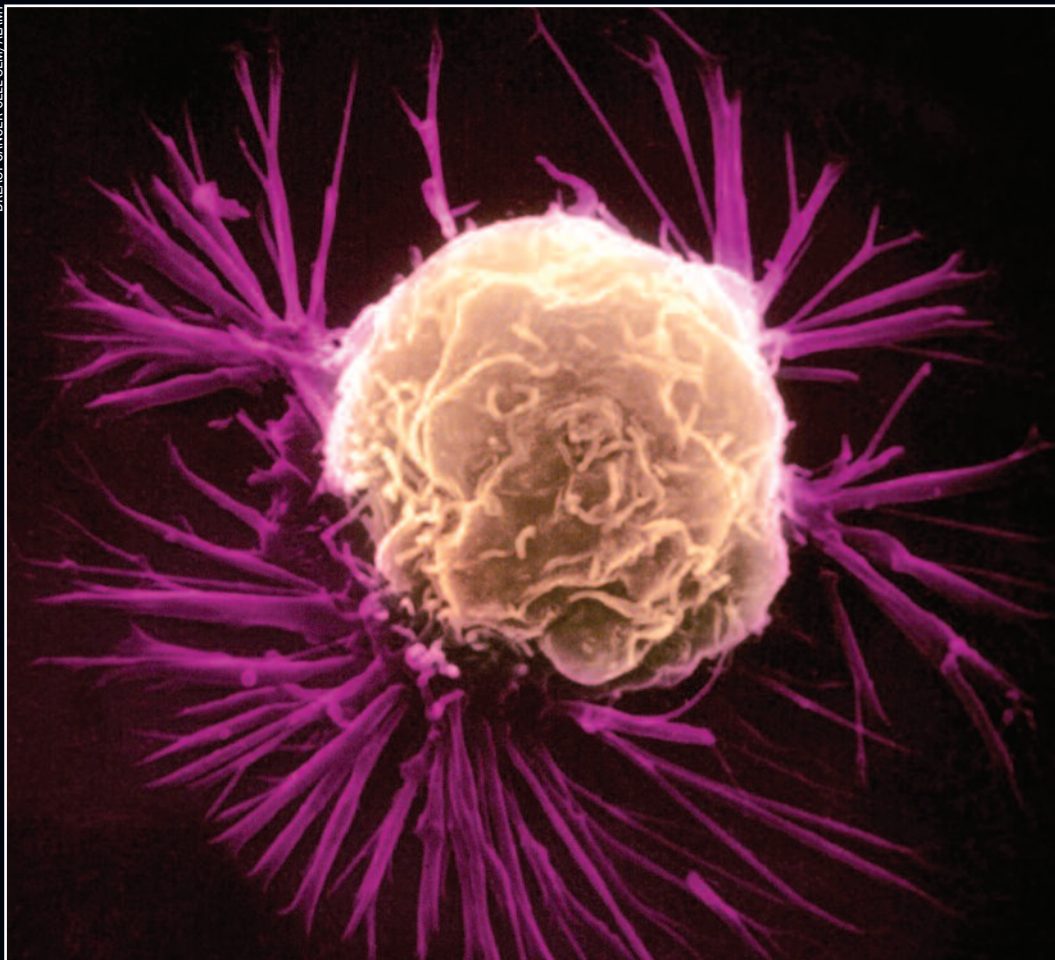
# Systems failure

A scandal involving clinical trials based on research that was riddled with errors shows that journals, institutions and individuals must raise their standards, argues

**Darrel Ince**







**C**hemotherapy is painful and debilitating. The side-effects include nausea, diarrhoea, extreme tiredness, loss of balance and loss of hair. All this happens while you wonder whether you will see your family and friends again.

A chemotherapeutic treatment that exacts less of a physical toll would benefit a great many people. In 2006, a group of researchers at Duke University announced in a research article a major breakthrough that promised precisely that. This was followed by several articles in the same vein; all were published in leading journals and had citation counts that any academic would envy. One paper, in the *New England Journal of Medicine*, was cited 290 times.

By 2009, three trials based on the research results were under way, with 109 cancer patients eventually enrolled. But the efforts never came to fruition – in fact, the trials were halted early, for the promise had been a hollow one. The research was riddled with major errors. This sad story has lessons for our universities, individual researchers and academic journals.

Over the past two years, the full story has emerged through the persistence of two biostatisticians, who established that the research was built on mud. The Duke University work was cutting edge and complex. It combined research into genomic data processing with the study of the effectiveness of medical therapies. It forms part of an important branch of medical research termed personalised medicine, in which an individual's genetic make-up is used to indicate an optimal therapy. The Duke researchers claimed that

genetic markers could predict the best course of drugs. Their work was considered so promising that clinical trials involving cancer patients were started.

When the first Duke papers on therapeutic regimes came to the attention of clinicians at the University of Texas MD Anderson Cancer Center in Houston, they were keen to try the techniques. Two of their biostatisticians, Keith Baggerly and Kevin Coombes, were asked to investigate. They discovered major problems with the statistics and the validity of the data, and pointed this out to the Duke researchers.

**By 2009, three trials based on the research results were under way, with 109 cancer patients eventually enrolled. The problem was that the research was riddled with errors**

Although some small errors were rectified, the researchers were adamant that the core work was valid.

Baggerly and Coombes then did what every good scientific researcher does: they issued short communications to the journals that had published the Duke work in which they pointed out the problems in the original research and in other related studies. A few of their criticisms were accepted, but some leading journals rejected their arguments.

Baggerly submitted three of those rejections to a recent US Institute of Medicine review into genomic data processing. One journal had concluded, based on a referee review, that the

issue centred on a statistical debate in which there was no right or wrong answer. Another journal, which had already published one letter in which Baggerly and Coombes set out criticisms, declined a second – largely because it had an editorial policy of not publishing multiple critiques of an article by the same authors. The third journal, which had published the Baggerly and Coombes' criticism of one paper, rejected that of a second paper without explanation despite the fact that the two researchers had asked for clarification.

By mid-2009, the questions and concerns had become more urgent as an academic problem had morphed into an ethical one. The frustrated biostatisticians discovered that cancer patients were already receiving treatments based on the research that they considered to be flawed. In the hope of reaching clinicians who were interested in using the Duke protocols, they sent a draft article to a biological journal. The feedback on that submission was that the piece was, in essence, too negative.

Their next move was to submit a paper to the reputable *Annals of Applied Statistics*, where they thought they would get a hearing. They set out all their criticisms and suggested that the Duke work might be putting patients at risk by directing therapy in the opposite way to that intended. It was accepted very quickly and published, in September 2009.

When staff at the US National Cancer Institute (NCI), one of the world's leading cancer research organisations, read the article, they contacted Duke University about some of the discrepancies Baggerly and Coombes had noted. The university then, to its credit,

launched a semi-independent review, involving two external reviewers as well as some of the university's senior managers. At the same time, the clinical trials were suspended.

What was not to the university's credit, however, was that it failed to pass on to the review panel one of the key Baggerly and Coombes communications.

The review panel was given the *Annals of Applied Statistics* paper, but not the biostatisticians' analysis of new data for two of the drugs used in the clinical trials, which had been issued while the review was under way. The new analysis claimed that all the Duke validation data were wrong. Baggerly and Coombes sent this information to the Duke managers who were overseeing the investigation. The managers forwarded it, via several intermediaries, to the two principal researchers, Anil Potti and Joseph Nevins, who were asked whether any of the criticisms were new.

**T**hese latest allegations were not, however, shown to the review team. According to the university's final report on the matter, this was because "Dr Nevins expressed his strong objection...believing that this was an improper intrusion by Dr Baggerly into an independent review process commissioned by the Duke Institutional Review Board" and because the claims amounted to nothing new. The review cleared the Duke team and gave permission for the clinical trials. Baggerly and Coombes were incredulous.

At the end of January 2010, the university announced that the clinical trials were restarting. In response, Baggerly and Coombes published their report that the Duke inquiry

had rejected.

The university at first refused to allow Baggerly and Coombes to see the external reviewers' report that was used to justify restarting the trials, claiming that it was confidential.

In May, however, following a freedom of information request, the biostatisticians obtained a copy from the NCI. When they read it, they concluded that it was an insufficient basis for restarting the trials. Their reaction was published in *The Cancer Letter*, a publication for researchers, clinicians and

**When Baggerly and Coombes read the external reviewers' report used to justify restarting the clinical trials, they concluded that it was an insufficient basis for proceeding**

staff in the pharmaceutical industry. Duke University did not respond.

On 16 July 2010, *The Cancer Letter* broke the story that Potti, one of the lead researchers, had falsified aspects of his curriculum vitae, including a claim to be a Rhodes scholar. This, at last, brought action.

Potti was placed on administrative leave, and Duke appears to have suspended the trials once again.

At the same time, a group of biostatisticians and bioinformaticians began to campaign for a pause in the trials and a closer examination of the research. Thirty-three leading researchers wrote to Harold Varmus, director of the NCI,

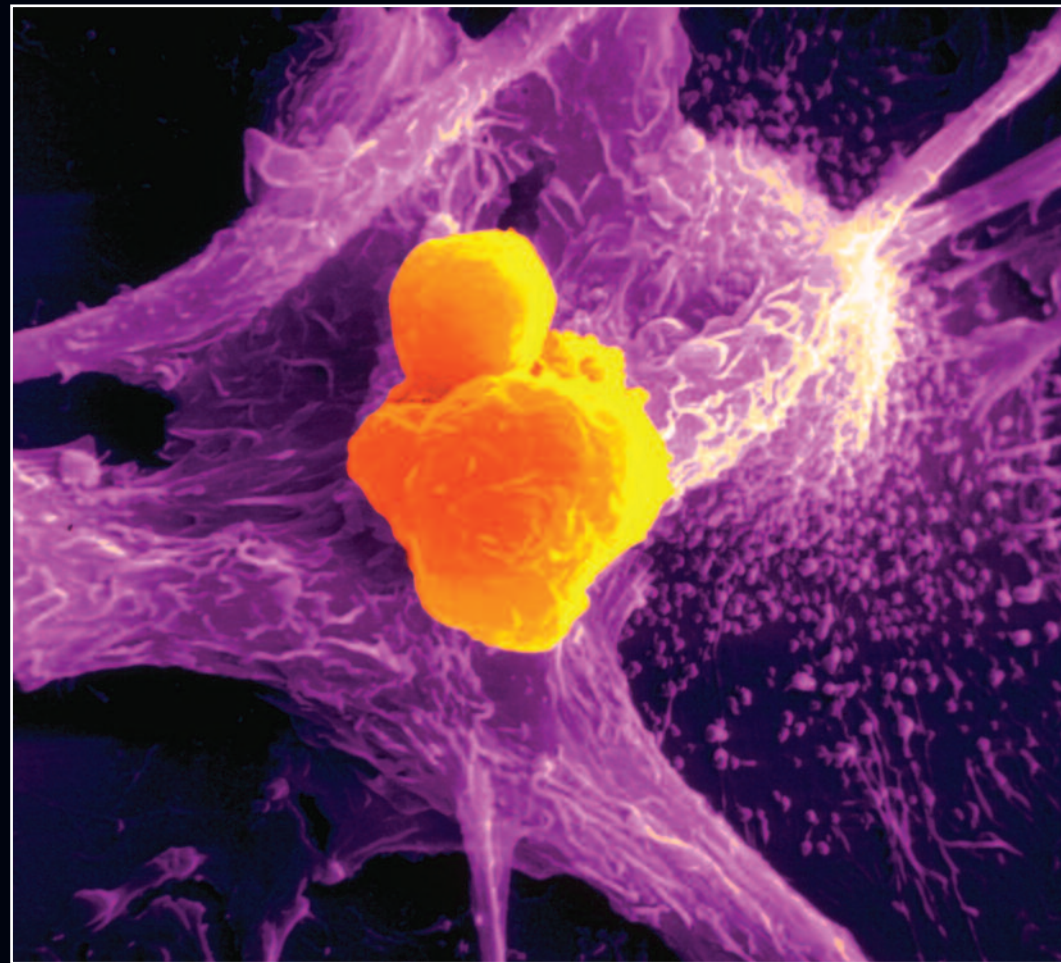
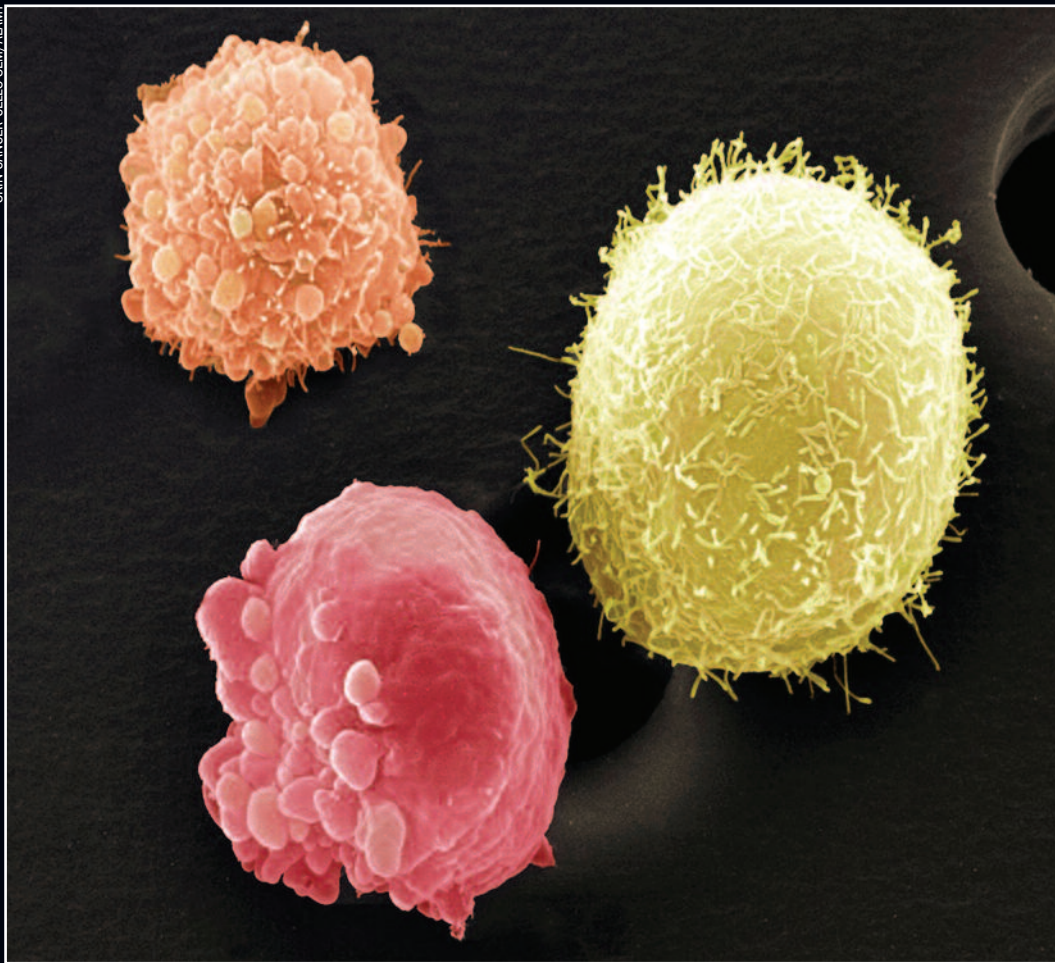
asking that the trials be suspended until the science was clarified. The trials were stopped and this time have not restarted.

Behind the scenes, the NCI, prompted by the *Annals of Applied Statistics* paper, had already begun to carry out some checks of its own. In April 2010, a reviewer noticed that an institute grant awarded to Potti included partial funding for a clinical trial, but none of the trials for which Potti was responsible acknowledged NCI support. This was important because the institute has direct legal cause for action only for those trials it supports. When Potti and Duke University were asked which trial the institute was helping to fund, it emerged that the chemotherapy study under scrutiny was the one. The NCI immediately asked for the raw data and code. In May, it told the university that it had been unable to reproduce the research results that were based on this material.

Although the institute had asked for the data and code to be provided quickly, any action that the NCI might then have taken was pre-empted by Duke's stopping the trials after the Potti CV revelations. The NCI's actions were made public at the end of last year when the organisation released several other reports detailing its work to the Institute of Medicine review, which began in late 2010.

So why had the Duke University review given the all-clear? The reason was that the external reviewers tasked with validating the research were working with corrupted databases. In the diplomatic words of the university's post-mortem report to the Institute of Medicine inquiry, the databases had "incorrect labelling...the samples also





appeared to be non-random and yielded robust predictions of drug response, while predictions with correct clinical annotation did not give accurate predictions”.

Potti admitted responsibility for the problems with the research and resigned. He had already been on administrative leave since the misstatements in his CV were discovered. Several of the Duke researchers’ papers were retracted, including those published in *Lancet Oncology*, the *Journal of Clinical Oncology* and *Nature Medicine*.

**N**o one comes out of this affair well apart from Baggerly and Coombes, *The Cancer Letter* and the *Annals of Applied Statistics*. The medical journals and the Duke researchers and senior managers should reflect on the damage caused. The events have blotted one of the most promising areas in medical research, harmed the reputation of medical researchers in general, blighted the careers of junior staff whose names are attached to the withdrawn papers, diverted other researchers into work that was wasted and harmed the reputation of Duke University.

What lessons should be learned from the scandal? The first concerns the journals. They were not incompetent. Their embarrassing lapses stemmed from two tenets shared by many journals that are now out of date in the age of the internet. The first is that a research paper is the prime indicant of research. That used to be the case when science was comparatively simple, but now masses of data and complex programs are used to establish results. The distinguished geophysicist Jon Claerbout has expressed this succinctly:

“An article about computational science in a scientific publication isn’t the scholarship itself, it’s merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions used to generate the figures.”

Baggerly and Coombes spent a long time trying to unravel the Duke research because they had only partial data and code. It should be a condition of publication that these are made publicly available.

The second tenet is that letters and discus-

**“The events have blotted a promising area of medical research, harmed the reputation of medical researchers, blighted the careers of junior staff and led to wasted work by researchers”**

sions about defects in a published paper announcing new research have low status. Journals must acknowledge that falsifiability lies at the heart of the scientific endeavour. Science philosopher Karl Popper said that a theory has authority only as long as no one has provided evidence that shows it to be deficient. It is not good enough for a journal to reject a paper simply because it believes it to be too negative.

Journals should treat scientists who provide contra-evidence in the same way that they treat those putting forward theories. For an amusing and anger-inducing account of how one researcher attempted to have published a

comment about research that contradicted his own work, see “How to Publish a Scientific Comment in 1 2 3 Easy Steps” (<http://scr.bi/4Mh3w>).

The second lesson is for universities. University investigations into possible research irregularities should be conducted according to quasi-legalistic standards. In his evidence to the Institute of Medicine inquiry, Baggerly stated that he and Coombes had been hindered by the incompleteness of the Duke review – specifically in that the university did not verify the provenance and accuracy of the data that the researchers supplied to the review, did not publish the review report, did not release the data that the external reviewers were given and withheld some of the information Baggerly and Coombes had provided to the review.

The university’s explanation for not passing on the new Baggerly and Coombes material was a “commitment to fairness to the faculty” and a senior member of the research team’s “conviction and arguments, and in recognition of his research stature”. A similar argument in a court of law would not have been allowed.

The third lesson is for scientists. When research involves data and computer software to process that data, it is usually a good idea to have a statistician on the team. At the “expense” of adding an extra name to a publication, statisticians provide a degree of validation not normally available from the most conscientious external referee. Indeed, the statistics used might merit an extra publication in an applied statistics journal. Statisticians are harsh numerical critics – that’s their job – but their involvement gives the

researcher huge confidence in the results. Currently the scientific literature, as evidenced by the major research journals, does not boast any great involvement by statisticians.

For all the talk about interdisciplinarity, there is often little cooperation in universities between researchers in different areas. Major scientific projects could benefit not only from input from statisticians, but also from computer scientists. In my early days as an external examiner, it was fairly common for a computing department to run a course on scientific computing, often at the behest of a science faculty. These have now disappeared. Some have been replaced by simple programming courses run by the computer services department. There is an opportunity here for an innovative university to create some really interesting courses.

The statistician Victoria Stodden is already running one such course at Columbia University, on reproducibility in science. Students are required to examine a piece of published research and try to reproduce the results. They are encouraged to critique the existing work, reapply the data analysis and document the work in a better way. If they discover problems, they are encouraged to publish their results.

**A** fourth lesson from the Duke affair concerns reproducibility. The components of a research article should be packaged and made readily available to other researchers. In the case of the Duke study, this should have included the program code and the data. This did not happen. Instead, Baggerly and Coombes spent about 200 days

exploring the partial materials provided to conduct their forensic investigation. In a pre-internet, pre-computer age, packaging-up was less of an issue. However, the past decade has seen major advances in scientific data-gathering technologies for which the only solution is the use of complex computer programs for analysis.

A number of tools are now being developing for packaging up research. Among the best is Sweave, a software system that combines an academic paper, the data described by the paper and the program code used to process

**“Duke implicitly acknowledges mistakes. It says ‘quantitative expertise is needed for complex analyses’ and ‘sustained statistical collaboration is critical to... management of complex datasets’”**

the data into an easily extractable form. There are also specific tools for genetic research, such as GenePattern, that have friendlier user interfaces than Sweave.

What is worrying is that more scandals will emerge, often as a result of the pressure on academics, who are increasingly judged solely on the volume of their publications (some systems even give an academic a numerical rating based on paper citation) and their grants, and on how patentable their work may be. Our universities are ill-prepared to prevent scandals happening or to cope with the after-effects when they do happen. There is a clash here between collegiality and the university as

a commercial entity that needs to be resolved.

In its official account to the Institute of Medicine inquiry – in effect a chronicle, a detailed description of the errors that were committed and a future agenda – Duke University implicitly acknowledges the mistakes. It states that “quantitative expertise is needed for complex analyses”, “sustained statistical collaboration is critical to assure proper management of these complex datasets for translation to clinical utility” and “the implementation and utilization of systems that provide the ability to track and record each step in these types of complex projects is critical”.

This document is to be recommended to academics and university managers alike, not just as background to this article, but also as a cautionary story and a source of action points and self-questioning for any university that prides itself on its scientific research and its ethical standards. It can be found at [www.cancerletter.com/categories/documents](http://www.cancerletter.com/categories/documents)

Despite the considerable technical detail, a non-geneticist will still be able to gain much from it. Many of us, myself included, have been sloppy in packaging up our research, but the intrusion of the computer and the internet, and the increasing commercial pressures on our universities demand higher standards.

**Darrel Ince is professor of computing at the Open University. He is writing an account of the Duke University affair with the title *The Cracks in Science*.**